

NBRP databases: databases of biological resources in Japan

Yukiko Yamazaki^{1,*}, Ryo Akashi², Yutaka Banno³, Takashi Endo⁴, Hiroshi Ezura⁵, Kaoru Fukami-Kobayashi⁶, Kazuo Inaba⁵, Tadashi Isa⁷, Katsuhiko Kamei⁸, Fumie Kasai⁹, Masatomo Kobayashi⁶, Nori Kurata¹, Makoto Kusaba¹⁰, Tetsuro Matuzawa⁴, Shohei Mitani¹¹, Taro Nakamura¹², Yukio Nakamura⁶, Norio Nakatsuji⁴, Kiyoshi Naruse¹³, Hironori Niki¹, Eiji Nitasaka³, Yuichi Obata⁶, Hitoshi Okamoto¹⁴, Moriya Okuma⁶, Kazuhiro Sato¹⁵, Tadao Serikawa⁴, Toshihiko Shiroishi¹, Hideaki Sugawara¹, Hideko Urushibara⁵, Masatoshi Yamamoto¹⁶, Yoshio Yaoita¹⁰, Atsushi Yoshiki⁶ and Yuji Kohara¹

¹National Institute of Genetics, ²University of Miyazaki, ³Kyushu University, ⁴Kyoto University, ⁵University of Tsukuba, ⁶BioResource Center, RIKEN, ⁷National Institute for Physiological Sciences, ⁸Chiba University, ⁹National Institute for Environmental Studies, ¹⁰Hiroshima University, ¹¹Tokyo Womens Medical University, ¹²Osaka City University, ¹³National Institute for Basic Biology, ¹⁴Brain Science Institute, RIKEN, ¹⁵Okayama University and ¹⁶Kyoto Institute of Technology, Japan

Received September 15, 2009; Revised October 15, 2009; Accepted October 16, 2009

ABSTRACT

The National BioResource Project (NBRP) is a Japanese project that aims to establish a system for collecting, preserving and providing bio-resources for use as experimental materials for life science research. It is promoted by 27 core resource facilities, each concerned with a particular group of organisms, and by one information center. The NBRP database is a product of this project. Thirty databases and an integrated database-retrieval system (BioResource World: BRW) have been created and made available through the NBRP home page (<http://www.nbrp.jp>). The 30 independent databases have individual features which directly reflect the data maintained by each resource facility. The BRW is designed for users who need to search across several resources without moving from one database to another. BRW provides access to a collection of 4.5-million records on bioresources including wild species, inbred lines, mutants, genetically engineered lines, DNA clones and so on. BRW supports summary browsing, keyword searching, and searching by DNA sequences or gene ontology. The results of searches provide links to online requests for distribution of research materials. A circulation system

allows users to submit details of papers published on research conducted using NBRP resources.

INTRODUCTION

Japanese bioresources for life science research have an 80 year history. Although some unique and precious collections of resources have been accumulated during this time, some of these have recently become dispersed and lost as a result of the increasing age of their providers or through shortages of funds, so that systems for providing useful collections of resources became inadequate. To improve this situation, it was necessary to establish a sustainable environment in which researchers could readily obtain bioresources, so the National BioResource Project (NBRP) began in 2002. The species included in the project were selected on the basis that they were indigenous to Japan or that they were model organisms currently studied or expected to be studied in the future by large numbers of researchers.

A major feature of this project is that it promotes the centralization of resources and information to ensure continuity. Resources are organized by species or groups of organisms, and a system of interaction between the resource and the information center [which belongs to the National Institute of Genetics (NIG)] was created to centralize the information. All the information in the databases is publicly available and several enhancements

*To whom correspondence should be addressed. Email: yyamazak@lab.nig.ac.jp

in the contents of the databases and in the services offered are being introduced.

Here, we outline some details of the BioResourceWorld (BRW) integrated database and several representative component databases. All the information described is available at <http://www.nbrp.jp/>.

NBRP Information Site is a gateway to access all the information of NBRP. The menu on the left-hand side beneath the heading 'Resource Center' provides links to all 30 individual databases, and gives the user direct access to the appropriate database for the organism of interest. The text box 'Keyword' near the upper center of the page is a query box for entering keyword searches of the BRW integrated database. Users can access the home page of BRW directly by clicking on 'Resource Integrated Search Site' in the menu bar near the top of the page. This menu bar also provides links to other pages: 'Journal' provides a link to the Research Resource Circulation (RRC) Web site for browsing and submission of papers reporting research using resources obtained through NBRP; 'Japan Genetic Resources' provides a link to a list of URLs for resources in Japan, but outside the NBRP group; and 'Worldwide Genetic Resources' provides a link to a list of URLs for bioresources worldwide. The total number of records currently available in the database for all bioresources and the subtotals for three individual classifications (animals, plants and microbes) are shown in the turquoise-colored box near the top left of the page.

NBRP DATABASES

An integrated NBRP database retrieval system: BioResourceWorld

BRW is an integrated database retrieval system that allows users to retrieve resources by using the body of information held by NBRP on a number of organisms. All the resources (~4.5 million items) are available for distribution. Keyword searching of BRW is directly available from the NBRP home page. A user performing a search can specify an organism and a resource category from pull-down menus. (The default values are 'All organisms' and 'All categories'.) When a user enters keywords and performs a search with the default settings, the total number of results is shown together with a breakdown of the number of strains and the number of DNA clones. This is followed by the list of resources broken down by organism. By clicking on a desired resource on the list, the user can access a detailed information page that provides links to the distribution-request site for each database (where this exists).

The BRW home page, which is linked from the top menu bar of the NBRP home page, contains the search functions described above and four tabs (ALL, DNA, BLAST and Gene Ontology). These four tabs provide, respectively, a summary display of all resources by organism groups, a summary display of DNA clones by organism groups, a BLAST search and a Gene Ontology (GO) search.

The BLAST search allows the user to search all organism groups and also to specify particular organisms or categories, such as genomic clones, cDNA clones, or libraries. The search results are shown as diagrams of alignments in which abbreviations of names of organisms or groups of organisms appear in a color-coded manner on the left-hand side of the diagram to help users recognize particular organisms. Individual sequences in hits have links to NCBI (1), DDBJ (2) and BRW, which gives the user access to the web sites from which the appropriate resources can be ordered.

The latest service of BRW is a search of resources on the basis of GO. The user can examine the hierarchical structure of a GO term and query it by using GO ID, GO term, or GO Gene. For example, if a user enters '000038' (or 'very-long-chain fatty acid metabolic process') into the GO-ID (or GO Term), a resource list containing 'mouse (5/5), drosophila (2/4), arabidopsis (10/10)' is shown. The numbers in parenthesis indicate the number of resources associated with this GO-ID (left) and its descendants (right). Thus this service allows users a more semantic search and provides them with a wider range of resources.

The GO search is currently at a testing stage, because not all resources contain GO information. The total number of resources mapped to a GO Term is displayed, and a list of relevant resources can be obtained by clicking the figures. Like other search modes, the GO allows the ordering of resources from search results. We plan to enhance the search of resources across organism by the addition of supporting bio-ontology data, such as phenotype, anatomy and development.

The information contained in BRW is updated concurrently with that in the individual databases.

Resource research circulation

If a researcher obtains good results with NBRP resources, such a result may be useful for later researchers who use the same resources. We therefore ask researchers to feed back information on papers that report the results of their studies, and we also collect other papers in which NBRP resources are used. This information is used to create an open database (RRC) of papers related to the NBRP resources. RRC also provides an online registration system through which a paper can be submitted merely by entering its PubMed identification number and the name of the resource used in the paper. We have asked many researchers to use this system to feed back on the papers that they have published. It would be ideal for experimental researchers if, as in the case of the accession numbers of the DNA Data Bank, we could establish a system in which detailed information on the bioresources described in 'Materials and Methods' sections of papers could be obtained from public databases and, furthermore, the materials themselves could be easily acquired.

The NBRP databases

Table 1 shows the names and features of the NBRP component databases, organized by organism group. As shown in Table 1, the NBRP resource collection covers

a wide range of taxa with indications of the presence or absence of the following collections: (i) naturally mated lines (including wild species, cultivated species, inbred lines and spontaneous mutants); (ii) genetically engineered organisms (including induced mutants, transgenic strains, transposon-inserted strains, deletion strains, consomic lines, genome-wide knockout strains, and enhancer trap lines); and (iii) DNA (plasmids, vectors and genomic/cDNA clones). The numbers of records of resources vary according to the group of organisms, and the resources in classification (3) account for 95% of the total. The External DB column in Table 1 indicates databases that have one-way links or cross links to external comprehensive databases for the model organisms. The Collaborating DB inside the parenthesis indicates an external database with which the NBRP database collaborates.

Types and names (identifiers) of the resources, and distribution/deposition methods (such as MTA) are essential information for all resources. For DNA clones, the accession numbers in the DNA Data Bank, together with homologous sequences, are the items of information common to all organism groups. NBRP also includes some activities in which the distribution of resources requires a prescribed review procedure (such as the Human ES-cells and Macaque). Activities where resources cannot be obtained by proliferation, but require non-invasive sampling or sampling from dead animals, are provided only for research that meets prescribed standards [such as the Great Ape Information Network (GAIN)]. The Global Biodiversity Information Facility (GBIF; <http://www.gbif.org/>) is an activity that forms the Japanese node of an international project that is involved in collecting and creating a database on specimens and observation data held in museums. Information from the GBIF and resources from the GAIN and Macaque databases are generally not distributed, so that they are not included in BRW.

As illustrative examples of the features described above, the databases for rice, *Escherichia coli*, *Caenorhabditis elegans* and the rat are described in more detail below, and some details of the RIKEN BioResource Center databases are also given.

NBRP-rice (Oryzabase) (25). Oryzabase, a comprehensive database that is currently used by many researchers, succeeded the Database of Resources and Trait Genes, which was launched in 1995. Genomic information has been added to the database on resources and genes so that it has become comprehensive and now functions as the NBRP database. The characteristic wild-rice collection consists of three different core collections. Core collection Rank 1 relates to 18 species from nine genomes and contains detailed phenotype data and many photographs. The collection of unique mutant strains includes strains for which trait genes have been identified, as well as strains that have phenotypes for which the responsible genes are not known. The former set of strains can be retrieved on the basis of genetic information, and the latter set of strains can be retrieved on the basis of the

phenotype. The data on both sets of strains include photographs.

We constantly update the dictionary of trait genes with data extracted from published papers, and we have also been promoting the establishment of correspondences between trait genes and the accession numbers of DNA sequences, ORF numbers of genome projects and displays of chromosomes by using physical maps. Oryzabase also provides access to a Web site for online submission of new rice genes, which is based on the Gene Nomenclature System for Rice as determined by the Committee on Gene Symbolization, Nomenclature and Linkage of the Rice Genetics Cooperative (CGSNL) (26). Users can give feedback on an individual gene through the detailed pages for that gene.

Basic information on rice, including the definitions of the tissue-specific developmental stages (such as embryo, inflorescence, leaf, root, anther, ovule, pollen mother cell, stoma or vascular bundle) and information on tissue-specific or developmental stage-specific gene expression are also available through this web site.

A version of Textpresso, a text-mining system for scientific papers developed by the Generic Model Organism Database and adapted for use with rice, has been constructed and is available through Oryzabase. It provides access to a total of ~20 000 rice-related papers (abstracts and titles) in PubMed, and is linked from the home page of Textpresso.

Oryzabase has links to two external databases: Gramene and IRRI. Genetic information is linked to the former database, and wild-strain information is linked to the latter. Oryzabase also allows BLAST searches to be made through all genomes or by chromosomes, provides a tool for extracting the specified region of genome sequences, and can provide downloadable text files of almost all information.

NBRP-E. coli. Of all the organisms in the NBRP collection, *E. coli* was the first to have its genome sequence determined. A feature of NBRP-*E. coli* is its genome-wide genetically engineered strain collections. At the web site of NBRP-*E. coli* (<http://www.shigen.nig.ac.jp/ecoli/strain/>), users can browse outlines and lists of collections and search full data or perform queries by specifying details. Although online requests are available, these often take time because of the complexity of the MTA process. For the convenience of users, a tracking system has been introduced that allows users to check the progress of their requests in real time.

The information center maintains profiling of *E. coli* chromosomes (PEC), an information site of a project on essential *E. coli* genes (27). Through PEC, the information center has made genome maps and genetic information available. The resource information contained in NBRP-*E. coli* is cross linked with PEC to permit access to resources through the maps. For example, by clicking the deletion regions of extensive deletion mutants, or by clicking the gene parts of mutants mutated by gene units, the user can access detailed pages that are linked to request pages for the resources. PEC has many links to external databases, including NCBI, UniProtKB, COG,

Table 1. The NBRP databases

Database name	Taxon	Wild/ inbred/ landrace	Mutants	Plasmid/ vector/ clones	Map (Physical or Linkage)	Gene (mutated genes/all genes)	Phenotype	SNPs	Blast service	Images	External DB (collaborating DB)[reference]	Featured contents/ services
1 NBRP-Macaque	Vertebrata-mammalia-primatae	0	0	0	P	M	0	0	0	0	MGJ[3], (IMSR, JMSR, http://www.shigen.nig.ac.jp/mouse/jmstr/)	graphical display of the phenotype data
2 NBRP-Mouse (Riken BRC)	Vertebrata-mammalia-murinae-mus	0	0	0	P	M	0	0	0	0	RGD[4], (JMSR)	
3 NBRP-Rat	Mammalia-murinae-rattus	0	0	0	P	M	0	0	0	0		
4 NBRP-Xenopus	Vertebrata-amphibia	0	0	0		M	0	0	0	0	ZFIN[5]	
5 NBRP-Zebrafish	Vertebrata-actinopterygii-cypriniformes-danio	0	0	0	P	M	0	0	0	0		
6 NBRP-Medaka	Vertebrata-actinopterygii-betuliniformes-oryzias	0	0	0		M	0	0	0	0	(Ghost,[6] CIPRO; http://cipro.bio.jp/new/)	atlas, phylogenetic tree
7 NBRP-Ciona	Ascidacea-ciona	0	0	0	P	A	0	0	0	0	FlyBase[7]	
8 NBRP-Drosophila (DGR)	Arthropoda-insecta-diptera-drosophilidae	0	0	0	P	A	0	0	0	0	Wormbase[8]	laeva period time
9 NBRP-Silkworm (SilkwormBase)	Arthropoda-insecta-lepidoptera-bombyx	0	0	0	L	M	0	0	0	0	Gramene[9], IRRI; http://beta.irri.org/index.php/Home/Welcome/	developmental stage
10 NBRP-C. elegans	Pseudocoelomata-nematoda	0	0	0	P,L	M	0	0	0	0	Frontpage.html	phenotype data
11 NBRP-Rice (Oryzabase)	Viridiplantae-poaceae-oryzaceae	0	0	0		A	0	0	0	0	(TriFLDB[10])	gene catalogue
12 NBRP-Barley	Viridiplantae-poaceae-triticeae-hordeum	0	0	0	L	A	0	0	0	0		
13 NBRP-Wheat (KOMUGI)	Viridiplantae-poaceae-triticeae	0	0	0	L	A	0	0	0	0		
14 NBRP-Arabidopsis (Riken BRC)	Viridiplantae-brassicales-ara bidopsis	0	0	0		M	0	0	0	0	(miyakogusa.jp[11], Soybean Full-length cDNA database[12])	
15 NBRP-Chrysanthemum	Viridiplantae-asterids-chrysanthemum	0	0	0		M	0	0	0	0	(KafT om; http://www.pgb.kazusa.or.jp/kaftom/), MIBASE[13])	
16 NBRP-Morning glory	Viridiplantae-solanales-ipomoeaceae	0	0	0	L	M	0	0	0	0		
17 NBRP-Lotus/Glycine (LegumeBase)	Fabales-fabaceae	0	0	0	L	M	0	0	0	0		
18 NBRP-Tomato	Viridiplantae-solanales-lycopersicon	0	0	0		M	0	0	0	0		
19 NBRP-Algae	14 phyla (eukaryota and bacteria)	0	0	0	P	A	0	0	0	0	SGD[14], geneDB[15] dictyBase[16]	photograph, phylogenetic tree
20 NBRP-Yeast	Fungi-ascomycota	0	0	0	P	A	0	0	0	0		
21 NBRP-Cellular slime mold	Mycetozoa-dictyostellida	0	0	0		A	0	0	0	0		
22 NBRP-Prokaryote E. coli	Bacteria-proteobacteria	0	0	0	P	A	0	0	0	0	PEC[17], EcoGene[18], EcoCyc[19], COG[20], NCBI[1], SwissProt[21], GTOP[22], KEGG[23], InterPro[24]	
23 NBRP-Prokaryote B. subtilis	Bacteria-firmicutes-bacillibacillaceae	0	0	0		M	0	0	0	0		
24 NBRP-Pathogenic microbes	Bacteria and protozoa	0	0	0		M	0	0	0	0		
25 NBRP-General microbes (Riken BRC, JCM)	Bacteria	0	0	0		M	0	0	0	0		

EcoCyc, GTOP, EcoGene and KEGG. We also perform similarity searches on gene sequences to add domain information from Pfam and PROSITE. In addition, PEC provides BLAST search through two different strains, MG1655 and W3110; all ORF, and essential genes. The tool for specifying the desired fragments of genome sequences is also available. Almost all the information in the database can be downloaded in the form of text files.

NBRP-C. elegans. *NBRP-C. elegans* (<http://www.shigen.nig.ac.jp/c.elegans/>) is a smaller database than the two databases discussed above. Each record consists only of the allele of the deletion strain, a systematic identification tag for the gene (the CGC name), information on the position of the gene on the chromosome, information on the positions of deletion regions, and primer information. However, the CGC name, allele and sequence are cross linked to the corresponding page of WormBase, a comprehensive database on *C. elegans*, providing an easy access to that database. In the case of this resource, requested mutants are isolated from the pool after the request. Users can check on the state of progress of isolation online. Mutants that have been isolated once will be listed as available mutants (isolated), so that other users can request the same mutants. *NBRP-C. elegans* is unique in that information on researchers who have received mutants from this project is made public online. Because almost all papers in which the resources are used contain the names of these resources, and the information is fed back from paper-registration sites, the *C. elegans* database automatically reflects information in the RRC. This is a model case of good circulation between resources and researchers.

NBRP-rat. Because the database of rat resources (<http://www.anim.med.kyoto-u.ac.jp/NBR/>) contains substantial characterization data from individual resources and has many tools for browsing this data, it is efficient in allowing researchers to find the resources best suited to their research from a range of trait information. For example, the top page displays a pie chart that shows a breakdown of the research fields in which the resources are used, and by clicking a research field of interest, the user can obtain a list of resources related to that particular research field. Another example, 'Phenome Project' provides 109 items of physiological, behavioral and anatomical phenome data in nine tables and in strain-distribution maps with the two items selected by the user as the abscissa and ordinate axes. Users can access detailed pages for strains by clicking data points in the maps. 'Genome' provides access to polymorphism data for 357 simple sequence-length polymorphism (SSLP) markers, obtained from investigations on more than 150 strains. A phylogenetic tree, constructed from polymorphism data, is also available, through which users can access detailed pages on resources.

Besides characterization data, the database also contains detailed pages on resources, including the preservation status, genetic status, research category, origin, genotype, references and a link to the Rat Genome Database (RGD). Recent enhancements include a BAC

browser for F344 and LE BAC end sequences, and the addition of functional polymorphism data obtained by comparing 16 disease-associated gene mutations among multiple strains. The ENU-induced mutant archive (28) is also provided at this site.

NBRP-RIKEN BRC

The RIKEN BioResource Center (BRC) was established in 2001 with the aim of becoming the finest core bioresource facility in the world. Since then, it has been engaged in collecting bioresources developed mainly by Japanese scientists. These bioresources include living strains of mice and *Arabidopsis*, human and animal cells, DNA materials and various microbes for which the RIKEN BRC has been designated the national core facility by the NBRP. The RIKEN BRC preserves these bioresources under conditions of strict quality control for provision to the scientific community. The RIKEN BRC also collects information on the whereabouts and characteristics of the bioresources, constructs databases, and offers the bioresource information to the research community within and outside Japan.

The Animal Search System (<http://www2.brc.riken.jp/lab/animal/search.php>) allows keyword searches to be performed on >2000 strains of mice available from the RIKEN BRC, including transgenic, knockout, inbred, wild-derived, ENU mutant and congenic strains. The system also provides detailed information on each strain, such as the strain name, description, gene details, references, availability status, health report, depositor, specific terms and conditions for distribution, and image(s), which capture the characteristics of the strain. Some gene symbols have links to the Mouse Genome Informatics (MGI) database, which leads to further detailed information on the particular gene. As well as being publicly available, the up-to-date information on mice is also sent to the International Mouse Strain Resource (IMSR, <http://www.findmice.org/>) database, to which the RIKEN BRC is a contributing repository. The RIKEN BRC Mouse Phenome Database (RMPD) (http://www.brc.riken.jp/rmpd/mouse_phenome_top.html) contains phenotypic data on the physiology, biochemistry, hematology and morphology of inbred, mutant, wild-derived and recombinant inbred strains of mice, and enables biomedical researchers to find appropriate strains for their researches.

The *Arabidopsis* transposon tagged lines can be searched in a web-based catalogue (<http://www.brc.riken.jp/lab/epd/catalog/transposon.html>), where information on insert positions of transposons and adjacent genes can be obtained for more than 15 000 lines. The SENDAI *Arabidopsis* Seed Stock Center (SASSC) database (<http://www.brc.riken.jp/lab/epd/SASSC/>) provides information on its collection (wild type and mutant), such as strain name, region of collection and phenotypic remarks. The RIKEN *Arabidopsis* full-length cDNA (RAFL) clone database (<http://www.brc.riken.jp/lab/epd/catalog/cdnaclone.html>) contains more than 250 000 clones and the users can retrieve clones by

NCBI accession number, AGI code, clone name or sequence homology. The Systematic Consolidation of *Arabidopsis* and other Botanical REsources (SABRE) database (<http://saber.epd.brc.riken.jp/sabre/SABRE0101.cgi>) offers searches of BRC plant DNA resources across the species that it contains.

The database of human and animal cells (<http://www2.brc.riken.jp/lab/cell/search.php>) has its origins in the database of the RIKEN Cell Bank, which began collecting and distributing cell resources in 1987. Starting from a stand-alone database for internal use, it has been transformed to allow searching of cell resources and the provision of the associated information through the Internet. Because the Cell Bank began its activity as a division of the BRC, its stock list has grown with the addition of various new kinds of cell resources, such as Epstein-Barr virus-transformed B cell lines, human somatic stem cells, embryonic stem cell lines and induced pluripotent stem cell lines. The function of the database has accordingly been enhanced to allow the presentation of other items of information that differ from resource to resource. The information available therefore depends on the resource; for example, information on the origin, morphology, culture conditions, restrictions on distribution, results of short tandem repeat analyses, images, etc. can be obtained for a conventional cell line resources.

Before the BRC was established, DNA materials were collected and distributed as an activity of the RIKEN DNA Bank. A keyword search system (<http://www.brc.riken.jp/lab/dna/search/index.html>) is available for DNA clones, vectors and recombinant adenoviruses. Resources are retrieved by plasmid name, by gene name or symbol, or by accession number. In the Geneset Bank database (<http://www.brc.riken.jp/lab/dna/en/GENESETBANK/index.html>), more than 20 illustrations for principal gene pathways are implemented, and DNA materials can be easily found according to the gene pathways to which they belong.

The microbial resource collection in RIKEN was founded in 1980 as the Japan Collection of Microorganisms (JCM), and the construction of a database for the collection began at its inception. The Web-based online catalog database (<http://www.jcm.riken.jp/JCM/catalogue.shtml>) was launched in 1995, and it has since been improved and updated. It now provides access to information on more than 11 000 available strains, which are searchable by their accession number, scientific name and keywords on strain data. It is also possible to search for JCM strains that are equivalent to those in other culture collections. The database contains various useful items of information about strains, such as culture media and conditions, history (including isolation source), taxonomic data and references. The DNA sequence data, critical for the phylogeny and the genome data are linked to the DDBJ database for each strain.

Future directions

We will continue to improve the content of individual databases and upgrade the functions available at the integrated database-retrieval site. We also hope to

expand external access to the databases and expand collaboration with other databases to permit access to our resources by a wider range of users. In particular, we will examine a new possibility for interconnecting reference data and resource databases with the aim of construction a virtual international network.

FUNDING

Ministry of Education, Culture, Sports, Science and Technology (MEXT) (to National Bio Resource Project). Funding for open access charge: The Management Expenses Grant for National University Cooperation, MEXT.

Conflict of interest statement. None declared.

REFERENCES

- Benson, A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J. and Sayers, E.W. (2009) GenBank. *Nucleic Acids Res.*, **37**, D26–D31.
- Sugawara, H., Ikeo, K., Fukuchi, S., Gojobori, T. and Tateno, Y. (2009) DDBJ dealing with mass data produced by the second generation sequencer. *Nucleic Acids Res.*, **37**, D16–D18.
- Blake, J.A., Bult, C.J., Eppig, J.T., Kadin, J.A., Richardson, J.E. and the Mouse Genome Database Group. (2009) The Mouse Genome Database genotypes:phenotypes. *Nucleic Acids Res.*, **37**, D712–D719.
- Dwinell, M.R., Worthey, E.A., Shimoyama, M., Bakir-Gungor, B., DePons, J., Laulederkind, S., Lowry, T., Nigram, R., Petri, V., Smith, J. *et al.* (2009) The Rat Genome Database 2009: variation, ontologies and pathways. *Nucleic Acids Res.*, **37**, D744–D749.
- Sprague, J., Bayraktaroglu, L., Bradford, Y., Conlin, T., Dunn, N., Fashena, D., Frazer, K., Haendel, M., Howe, D.G., Knight, J. *et al.* (2008) The Zebrafish Information Network: the zebrafish model organism database provides expanded support for genotypes and phenotypes. *Nucleic Acids Res.*, **36**, D768–D772.
- Satou, Y., Kawashima, T., Shoguchi, E., Nakayama, A. and Satoh, N. (2005) An integrated database of the ascidian, *Ciona intestinalis*: towards functional genomics. *Zool. J. Linn. Soc.*, **147**, 837–843.
- Tweedie, S., Ashburner, M., Falls, K., Leyland, P., McQuilton, P., Marygold, S., Millburn, G., Osumi-Sutherland, D., Schroeder, A., Seal, R. *et al.* (2009) FlyBase: enhancing *Drosophila* Gene Ontology annotations. *Nucleic Acids Res.*, **37**, D555–D559.
- Rogers, A., Antoshechkin, I., Bieri, T., Blasiar, D., Bastiani, C., Canaran, P., Chan, J., Chen, W.J., Davis, P., Fernandes, J. *et al.* (2008) WormBase2007. *Nucleic Acids Res.*, **36**, D612–D617.
- Liang, C., Jaiswal, P., Hebbard, C., Avraham, S., Buckler, E.S., Casstevens, T., Hurwitz, B., McCouch, S., Ni, J., Pujar, A. *et al.* (2008) Gramene: a growing plant comparative genomics resource. *Nucleic Acids Res.*, **36**, D947–D953.
- Mochida, K., Yoshida, T., Sakurai, T., Ogihara, Y. and Shinzaki, K. (2009) TriFLDB: a database of clustered full-length coding sequences from Triticeae with applications to comparative grass genomics. *Plant Physiol.*, **150**, 1135–1146.
- Sato, S., Nakamura, Y., Kaneko, T., Asamizu, E., Kato, T., Nakao, M., Sasamoto, S., Watanabe, A., Ono, A., Kawashima, K. *et al.* (2008) Genome structure of the legume, *Lotus japonicus*. *DNA Res.*, **15**, 227–239.
- Umezawa, T., Sakurai, T., Totoki, Y., Toyoda, A., Seki, M., Ishiwata, A., Akiyama, K., Kurotani, A., Yoshida, T., Mochida, K. *et al.* (2008) Sequencing and analysis of approximately 40,000 soybean cDNA clones from a full-length-enriched cDNA library. *DNA Res.*, **15**, 333–346.
- Yano, K., Watanabe, M., Yamamoto, N., Tsugane, T., Aoki, K., Sakurai, N. and Shibata, D. (2006) MiBASE: a database of a miniature tomato cultivar Micro-Tom. *Plant Biotechnol.*, **23**, 195–198.
- Hong, E.L., Balakrishnan, R., Dong, Q., Christie, K.R., Park, J., Binkley, G., Costanzo, M.C., Dwight, S.S., Engel, S.R., Fisk, D.G. *et al.* (2008) Gene Ontology annotations at SGD: new data

- sources and annotation methods. *Nucleic Acids Res.*, **36**, D577–D581.
15. Hertz-Fowler, C., Peacock, C.S., Wood, V., Aslett, M., Kerhornou, A., Mooney, P., Tivey, A., Berriman, M., Hall, N., Rutherford, K. *et al.* (2004) GeneDB: a resource for prokaryotic and eukaryotic organisms. *Nucleic Acids Res.*, **32**, D339–D343.
 16. Fey, P., Gaudet, P., Curk, T., Zupan, B., Just, E.M., Basu, S., Merchant, S.N., Bushmanova, Y.A., Shaulsky, G., Kibbe, W.A. *et al.* (2009) dictyBase—a Dictyostelium bioinformatics resource update. *Nucleic Acids Res.*, **37**, D515–D519.
 17. Yamazaki, Y., Niki, H. and Kato, J. (2008) Profiling of *Escherichia coli* Chromosome Database. *Methods Mol. Biol.*, **416**, 385–389.
 18. Rudd, K.E. (2000) EcoGene: a genome sequence database for *Escherichia coli* K-12. *Nucleic Acids Res.*, **28**, 60–64.
 19. Keseler, I.M., Bonavides-Martínez, C., Collado-Vides, J., Gama-Castro, S., Gunsalus, R.P., Johnson, D.A., Krummenacker, M., Nolan, L.M., Paley, S., Paulsen, I.T. *et al.* (2009) EcoCyc: a comprehensive view of *Escherichia coli* biology. *Nucleic Acids Res.*, **37**, D464–D470.
 20. Tatusov, R.L., Fedorova, N.D., Jackson, J.D., Jacobs, A.R., Kiryutin, B., Koonin, E.V., Krylov, D.M., Mazumder, R., Mekhedov, S.I., Nikolskaya, A.N. *et al.* (2003) The COG database: an updated version includes eukaryotes. *BMC Bioinformatics*, **4**, 41.
 21. UniProt Consortium. (2009) The Universal Protein Resource (UniProt) 2009. *Nucleic Acids Res.*, **37**, D169–D174.
 22. Fukuchi, S., Homma, K., Sakamoto, S., Sugawara, H., Tateno, Y., Gojobori, T. and Nishikawa, K. (2009) The GTOP database in 2009: updated content and novel features to expand and deepen insights into protein structures and functions. *Nucleic Acids Res.*, **37**, D333–D337.
 23. Kanehisa, M., Ataki, M., Goto, S., Hattori, M., Hirakawa, M., Itoh, M., Katayama, T., Kawashima, S., Okuda, S., Tokimatsu, T. *et al.* (2008) KEGG for linking genomes to life and the environment. *Nucleic Acids Res.*, **36**, D480–D484.
 24. Hunter, S., Apweiler, R., Attwood, T.K., Bairoch, A., Bateman, A., Binns, D., Bork, P., Das, U., Daugherty, L., Duquenne, L. *et al.* (2009) InterPro: the integrative protein signature database (2009). *Nucleic Acids Res.*, **37**, D224–D228.
 25. Yamazaki, Y. and Kurata, N. (2006) Oryzabase, An Integrated Biological and Genome Information Database for Rice. *Plant Physiol.*, **140**, 12–17.
 26. Susan, R.M. and CGSNL (Committee on Gene Symbolization, Nomenclature and Linkage, Rice Genetics Cooperative. (2008) Gene nomenclature system for rice. *Rice*, **1**, 72–84.
 27. Kato, J. and Hashimoto, M. (2008) Construction of long chromosomal deletion mutants of *Escherichia coli* and minimization of the genome. *Methods Mol. Biol.*, 279–293.
 28. Mashimo, T., Yanagihara, K., Tokuda, S., Voigt, B., Takizawa, A., Nakajima, R., Kato, M., Hirabayashi, M., Kuramoto, T. and Serikawa, T. (2008) An ENU-induced mutant archive for gene targeting in rats. *Nat. Genet.*, **40**, 514–515.